SEQUENCE LISTING

<110> Kent, Stephen B.H.
Muir, Tom W.
Dawson, Philip E.

<120> SYNTHESIS OF PROTEIN BY NATIVE CHEMICAL LIGATION

<130> gry0030p

<140> 08/945,997
<141> 1998-02-12

<150> PCT/US95/05668
<151> 1995-05-04

<160> 20

<170> PatentIn Ver. 2.1

<210> 1
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

Okay to enter JR 8.19.2002

<220>
<221> SITE
<222> (5)
<223> wherein COSH is thioacid

<400> 1
Leu Tyr Arg Ala Gly
1 5

<210> 2
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<400> 2
Cys Arg Ala Glu Tyr Ser
1 5

<210> 3
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<220>
<221> SITE
<222> (5)
<223> wherein COSBn is benzyl thioester

<400> 3
Leu Tyr Arg Ala Gly
1 5

<210> 4
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<220>
<221> SITE
<222> (5)
<223> wherein Gly is modified and represented by Gly-alphaCOS-CH2C(NHAc)CO2H

<400> 4
Leu Tyr Arg Ala Gly
1 5

<210> 5
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<400> 5
Leu Tyr Arg Ala Gly Cys Arg Ala Glu Tyr Ser
1 5 10

<210> 6
<211> 5

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: peptide

<220>
<221> SITE
<222> (5)
<223> wherein SCH2COOH is 2-thioacetic acid

<400> 6
Leu Tyr Arg Ala Gly
1 5

<210> 7
<211> 33
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (33)
<223> wherein COSH is thioacid

<220>
<221> SITE
<222> (1)
<223> wherein Msc is 2-methyl-sulfonyl-ethyloxy-carbonyl

<400> 7
Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
1 5 10 15

Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
20 25 30

Ala

<210> 8
<211> 33
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (33)
<223> wherein COSBn is benzyl thioester

<400> 8
Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
1 5 10 15

Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
20 25 30

Ala

<210> 9
<211> 39
<212> PRT
<213> Homo sapiens

<400> 9
Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu Leu
1 5 10 15

Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys Phe
20 25 30

Leu Lys Arg Ala Glu Asn Ser
35

<210> 10
<211> 72
<212> PRT
<213> Homo sapiens

<220>
<221> SITE
<222> (72)
<223> SH4

<400> 10
Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
1 5 10 15

Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
20 25 30

Ala Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu
35 40 45

Leu Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys
50 55 60

Phe Leu Lys Arg Ala Glu Asn Ser
65 70

<210> 11
<211> 40
<212> PRT
<213> Human immunodeficiency virus

<220>
<221> SITE
<222> (40)
<223> wherein COSNB is 5-thio-2-nitro-benzoic acid ester

<220>
<221> SITE
<222> (27)

<400> 11
Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

Ile Glu Glu Met Asn Leu Pro Gly
35 40

<210> 12
<211> 59
<212> PRT
<213> Human immunodeficiency virus

<220>
<221> SITE
<222> (27)
<223> wherein Xaa is 2-Aminobutyric acid

<220>
<221> SITE
<222> (55)
<223> wherein Xaa is 2-Aminobutyric acid

<400> 12
Cys Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val
1 5 10 15

Arg Gln Tyr Asp Gln Ile Pro Val Glu Ile Xaa Gly His Lys Ala Ile
20 25 30

Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn
35 40 45

Leu Leu Thr Gln Ile Gly Xaa Thr Leu Asn Phe
50 55

<210> 13
<211> 40
<212> PRT
<213> Human immunodeficiency virus

<220>
<221> SITE
<222> (40)
<223> wherein COSBn is ??

<220>
<221> SITE
<222> (40)
<223> wherein COSBn is benzyl thio ester

<400> 13
Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

Ile Glu Glu Met Asn Leu Pro Gly
35 40

<210> 14
<211> 40
<212> PRT
<213> Human immunodeficiency virus

<220>
<221> SITE
<222> (40)
<223> wherein COSPh is phenyl thioester

<400> 14
Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

Ile Glu Glu Met Asn Leu Pro Gly
35 40

<210> 15
<211> 99
<212> PRT
<213> Human immunodeficiency virus

<220>
<221> SITE
<222> (67)
<223> wherein Xaa is amino butyric acid

<220>
<221> SITE
<222> (95)
<223> wherein Xaa is 2-Aminobutyric acid

<400> 15
Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
20 25 30

Ile Glu Glu Met Asn Leu Pro Gly Cys Trp Lys Pro Lys Met Ile Gly
35 40 45

Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
50 55 60

Glu Ile Xaa Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65 70 75 80

Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Xaa Thr
85 90 95

Leu Asn Phe

<210> 16
<211> 48
<212> PRT
<213> Bacillus amyloliquefaciens

<220>
<221> SITE
<222> (48)
<223> wherein COSNB is 5-thio-2-nitro benzoic acid ester

<400> 16
Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr
1 5 10 15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala
20 25 30

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly
35 40 45

<210> 17
<211> 62
<212> PRT
<213> Bacillus amyloliquefaciens

<400> 17
Cys Ser Ile Gly Gly Asp Ile Phe Ser Asn Arg Glu Gly Lys Leu Pro
1 5 10 15

Gly Lys Ser Gly Arg Thr Trp Arg Glu Ala Asp Ile Asn Tyr Thr Ser
20 25 30

Gly Phe Arg Asn Ser Asp Arg Ile Leu Tyr Ser Ser Asp Trp Leu Ile
35 40 45

Tyr Lys Thr Thr Asp His Tyr Gln Thr Phe Thr Lys Ile Arg
50 55 60

<210> 18
<211> 48
<212> PRT
<213> Bacillus amyloliquefaciens

<220>
<221> SITE
<222> (48)
<223> wherein COSBn is benzyl thio ester

<400> 18
Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr
1 5 10 15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala
20 25 30

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly
35 40 45

<210> 19
<211> 48
<212> PRT
<213> Bacillus amyloliquefaciens

<220>
<221> SITE
<222> (48)
<223> wherein COSPh is phenyl thio ester

<400> 19

```
Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr
  1               5                  10                  15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala
             20                  25                  30

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly
         35                  40                  45
```

<210> 20
<211> 110
<212> PRT
<213> Bacillus amyloliquefaciens

<400> 20

```
Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr
  1               5                  10                  15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala
             20                  25                  30

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly
         35                  40                  45

Cys Ser Ile Gly Gly Asp Ile Phe Ser Asn Arg Glu Gly Lys Leu Pro
     50                  55                  60

Gly Lys Ser Gly Arg Thr Trp Arg Glu Ala Asp Ile Asn Tyr Thr Ser
 65                  70                  75                  80

Gly Phe Arg Asn Ser Asp Arg Ile Leu Tyr Ser Ser Asp Trp Leu Ile
                 85                  90                  95

Tyr Lys Thr Thr Asp His Tyr Gln Thr Phe Thr Lys Ile Arg
            100                 105                 110
```